IN THE SPECIFICATION:

Please amend paragraph number **[0005]** as follows:

**[0005]** The memory bit transistor's gate terminal connects to a wordline (towline) 16. The wordline, which connects to a multitude of memory bits, consists of an extended segment of the same polysilicon used to form the transistor's gate. The wordline is physically orthogonal to the digitline. A memory array, shown in FIG. 2, is created by tiling a selected quantity of memory bits together such that memory bits along a given digitline do not share a common wordline and such that memory bits along a common wordline do not share a common digitline. FIG. 3 contains an example of a memory array formed by tiling memory bits. There are several features of this layout that need illumination. First, note that the memory bits are in pairs to permit the sharing of a common contact to the digitline. This feature reduces the array size by eliminating unnecessary duplication. Second, note that any given wordline only forms (crosses) a memory bit transistor on alternating digitlines. This feature allows the formation of digitline pairs and ensures that wordline activation enables transistors only on alternate digitlines. Digitline pairs are an inherent feature in folded digitline arrays, as depicted in FIG. 3. An alternate array structure called open digitline architecture can also be used. A thorough understanding of both folded and open architectures by those skilled in the art assists in appreciating the characteristics and benefits of the bi-level digitline of the present invention. The open digitline and folded digitline architectures both have distinct advantages and disadvantages. While open digitline architectures achieve smaller array layouts by virtue of using smaller $6F^2$ memory bit cells, they also suffer from poor noise performance. A relaxed wordline pitch which stems from the $6F^2$ memory bit simplifies the task of wordline driver layout. Sense amplifier layout, though, is difficult because the array configuration is inherently half pitch -- one sense amplifier for every two digitlines. Folded digitline architectures, on the other hand, have superior ~~signal to noise,~~ signal-to-noise, at the expense of larger, less efficient array layout. Good ~~signal to noise~~ signal-to-noise performance stems from the adjacency of true and complement digitlines and the capability to twist these digitline pairs. For example, U.S. Patent No. 5,107,459 to Chu et al., issued April 21, 1992 describes a stacked digitline architecture

which uses lateral and vertical twisting. This technique, however, allows differential noise to be experienced on the digitlines which creates difficulty for differential sense amplifiers. Sense amplifier layout in the folded digitline architecture is simplified because the array configuration is quarter pitch -- one sense amplifier for every four digitlines. Wordline driver layout is more difficult since the wordline pitch is effectively reduced in folded architectures.

Please amend paragraph number **[0006]** as follows:

**[0006]** For the reasons stated above, and for other reasons stated below which will become apparent to those skilled in the art upon reading and understanding the present specification, there is a need in the art for a new array architecture which combines the advantages of both folded and open digitline architectures while avoiding their respective disadvantages. To meet this objective, the architecture needs to include the following features and characteristics: an open digitline memory bit configuration, a small $6F^2$ memory bit, and a small, efficient array layout. The memory must also include a folded digitline sense amplifier configuration, adjacent true and complement digitlines, and twisted digitline pairs to achieve a high ~~signal to noise~~ signal-to-noise ratio. Further, a relaxed wordline pitch should be used to facilitate better layout.

Please amend paragraph number **[0007]** as follows:

**[0007]** The above-mentioned problems with digitline architectures and other problems are addressed by the present invention and will be understood by reading and studying the following specification. A memory device is described which reduces overall die size beyond that obtainable from either the folded or open digitline architectures. A ~~signal to noise~~ signal-to-noise performance is achieved which approaches that of the folded digitline architecture.

Please amend paragraph number **[0063]** as follows:

**[0063]** After the cell access is complete, the sensing operation can commence. The reason for forming a digitline pair will now become apparent. FIG. 5 contains a schematic diagram for a simplified sense amplifier circuit. Note that it consists of a cross-coupled PMOS pair of P-sense amplifiers 18 and a cross-coupled NMOS pair of N-sense amplifiers 20. The NMOS pair or N-sense-amp common node is labeled NLAT* (for N-sense-amp LATch) in FIG. 5. Similarly, the P-sense-amp common node is labeled ACT (for ACTive pull-up). Initially, NLAT* is biased to Vcc/2 and ACT is biased to VSS or ground. Since the digitline pair D0 and D0* are both initially at Vcc/2 volts, the N-sense-amp transistors remain off due to zero Vgs potential. Similarly, both P-sense-amp transistors remain off due to their negative Vgs potential. As discussed in the preceding paragraph, a signal voltage develops between the digitline pair when the memory bit access occurs. While one digitline contains charge from the cell access, the other digitline serves as a reference for the sensing operation. The sense amplifier firing generally occurs sequentially rather than concurrently. The N-sense-amp fires first and the P-sense-amp second. The N-sense-amp is generally a better amplifier than the P-sense-amp because of the higher drive of NMOS transistors and better $V_{th}$ matching. This provides for better sensing characteristics and lower probability of errors. FIG. 6 contains waveforms for the sensing operation. Dropping the NLAT* signal toward ground will fire the N-sense-amp. As the voltage between NLAT* and the digitlines approaches $V_{th}$, the NMOS transistor, whose gate connection is to the higher voltage digitline, will begin to conduct. Transistor conduction occurs first in the subthreshold region, progressing to the saturation region as the gate to source voltage exceeds $V_{th}$. Conduction results in the discharge of the low voltage digitline toward the NLAT* voltage. Ultimately, NLAT* will reach ground, bringing the digitline with it. Note that the other NMOS transistor will not conduct since its gate voltage derives from the low voltage digitline, which is discharging toward ground. In reality, parasitic coupling between the digitlines and limited subthreshold conduction by the second transistor will reduce the high digitline voltage.

Please amend paragraph number **[0066]** as follows:

**[0066]** Memory Cells and ~~Arrays.~~ Arrays: The primary advantage of DRAM, over other types of memory technology, is low cost. This advantage arises from the simplicity and scaling characteristics of its 1T1C memory cell. Although the DRAM memory bit encompasses simple concepts, its actual design and implementation are highly complex. Successful, cost-effective DRAM designs require a tremendous amount of process technology.

Please amend paragraph number **[0076]** as follows:

**[0076]** Sense Amplifier ~~Elements.~~ Elements: The term "sense amplifier" refers to a collection of circuit elements that pitch up to the digitlines of a DRAM array. This collection most generally includes isolation transistors, devices for digitline equilibration and bias, one or more N-sense amplifiers, one or more P-sense amplifiers, and devices to connect selected digitlines to I/O signal lines. All of these circuits along with the wordline driver circuits, to be discussed later, are called pitch cells. This designation comes from the requirement that the physical layout for these circuits is constrained by the digitline and wordline pitches of an array of memory bits. For example, the sense amplifier layout for a specific digitline pair (column) generally consumes the space of four digitlines. This is commonly referred to as quarter-pitch or four-pitch, such that one sense amplifier exists for every four digitlines.

Please amend paragraph number **[0084]** as follows:

**[0084]** FIG. 23 shows a sense amplifier block commonly utilized in double or triple metal designs. It features two P-sense amplifiers 18 placed outside the isolation transistors 38, a pair of EQ/Bias devices as a gate terminal 36, a single N-sense amplifier 20, and a single I/O pass transistor 17 for each digitline. This design is quarter pitch, as are FIGs. 23 and 24, since only half of the sense amplifiers required for an array are on either side. Placement of the P-sense amplifiers outside the isolation devices is necessary since a full one level (Vcc) cannot pass through NMOS ISO transistors whose gate terminals are driven to Vcc. EQ/Bias devices or gate terminals 36 are also placed outside the ISO devices to permit continued equilibration of

digitlines in the isolated arrays. The I/O transistor gate terminals for four adjacent digitlines connect to a common CSEL signal. Each of the four I/O transistors is tied to a separate I/O bus. This sense amplifier, although simple to implement, is somewhat larger than other designs due to the presence of two P-sense amplifiers.

Please amend paragraph number **[0086]** as follows:

**[0086]** The sense amplifier block shown in FIG. 25 is at or near the minimum configuration. This design features single N-sense amplifiers 20 and P-sense amplifiers 18 ~~amplifiers~~ placed between the isolation transistors. To write full logic levels requires that the isolation transistors be either depletion mode devices or that the gate voltages be boosted above Vcc by at least one $V_{th}$. This design still includes a pair of EQ/Bias circuits to maintain equilibration on isolated arrays. Only a few designs have ever tried to operate with a single EQ/Bias circuit residing within the isolation devices. Most designers consider floating digitlines a risky proposition since cell leakage increases and sensing operations degrade as the digitlines drift away from the Vcc/2 precharge level. Future DRAM designs implemented on SOI (Silicon On Insulator) could permit the use of single EQ/Bias circuits, though, since this technology has negligible junction leakage.

Please amend paragraph number **[0089]** as follows:

**[0089]** Row Decoder ~~Elements.~~ Elements: Row decode circuits are similar to sense amplifier circuits in that they also pitch up to memory bit arrays and have a variety of implementations. A row decode block consists of two basic elements, a wordline driver and an address decoder tree. There are three basic configurations for wordline driver circuits that include the NOR driver, the inverter (CMOS) driver, and the bootstrap driver. Additionally, the drivers and associated decode trees can either be configured as local row decodes for each array section or as global row decodes which drive a multitude of array sections. Global row decodes connect to multiple arrays through metal wordline straps. The straps are stitched to the polysilicon wordlines at specific intervals dictated by the polysilicon resistance and the desired

RC wordline time constant. Most processes that strap wordlines with metal do not silicide the polysilicon, although doing so would reduce the number of stitch regions required. Strapping wordlines and using global row decoders obviously reduce die size -- in some cases very dramatically. The penalty to strapping is that it requires an additional metal layer and that this layer is at minimum array pitch. This puts a tremendous burden on process technologists in which three conductors are at minimum pitch -- wordlines, digitlines, and wordline straps. Distributed row decoders, on the other hand, do not require metal straps, but do require additional die size. It is highly advantageous to reduce the polysilicon resistance in order to stretch the wordline length and reduce the number of needed row decodes especially on large DRAMs such as the 1gigabit.

Please amend paragraph number **[0094]** as follows:

**[0094]** Address decode trees are the final element of the row decode block to be discussed. Decode trees are constructed from all types of logic -- static, dynamic, pass gate, or a combination thereof. Regardless of what type of logic ~~that~~ an address decoder is implemented with, the layout must completely reside beneath the row address signal lines to constitute an efficient, minimal design. In other words, the metal address tracks dictate the die area available for the decoder. For DRAM designs that utilize global row decode schemes, the penalty for inefficient design may be insignificant, but for distributed local row decode schemes, the die area penalty can be significant. As with memory bits and sense amplifiers, any time invested in row decode optimization is well spent.

Please amend paragraph number **[0099]** as follows:

**[0099]** Architectural ~~Characteristics.~~ Characteristics: A detailed description of the two most prevalent array architectures under consideration for future large scale DRAMs is provided -- the aforementioned open digitline and folded digitline architectures. To provide a viable point for comparison, each architecture will be employed in the theoretical construction of 32 Mbit memory blocks for use in a 256 Mbit DRAM. Design parameters and layout rules from

a typical 0.25$\mu$m DRAM process provide the needed dimensions and constraints for the analysis. Some of these parameters are shown in Table 2. Examination of DRAM architectures in the light of a real world design problem permits a more objective and unbiased comparison. An added benefit to this approach is that the strengths and weaknesses of either architecture should become readily apparent.

Please amend paragraph number **[0100]** as follows:

**[0100]** Open Digitline Array ~~Architecture.~~ Architecture: The open digitline array architecture was the prevalent architecture prior to the 64Kbit DRAM. A modern embodiment of this architecture as shown in FIG. 35 is constructed with multiple cross-point array cores separated by strips of sense amplifier blocks in one axis and either row decode blocks or wordline stitching regions in the other axis. Each 128Kbit array core is built using $6F^2$ memory bit cell pairs. There are a total of 131,072 ($2^{17}$) functionally addressable memory bits arranged in 264 rows and 524 digitlines. The 264 rows consist of 256 actual wordlines, 4 redundant wordlines and 4 dummy wordlines. The 524 digitlines consist of 512 actual digitlines, 8 redundant digitlines, and 4 dummy digitlines. Photolithography problems usually occur at the edge of large repetitive structures such as memory bit arrays. These problems produce malformed or nonuniform structures, rendering the edge cells useless. Inclusion of dummy memory bits, wordlines and digitlines on each array edge ensures that these problems only occur on dummy cells, leaving live cells unaffected. Although dummy structures make each array core larger, they also significantly improve device yield, making them necessary items on all DRAM designs.

Please amend paragraph number **[0102]** as follows:

**[0102]** Wordline ~~length,~~ length is limited by the maximum allowable RC time constant of the wordline. To ensure acceptable access time for the 256 Mbit DRAM, the wordline time constant should be kept below four nanoseconds. For a wordline connected to N memory bits, the total resistance and capacitance using 0.3$\mu$m polysilicon are $Rwl = Rs \cdot N \cdot Pwl \div 0.3\mu m$

ohms and $Cwl = Cw6 \cdot N$ Farads, respectively. Table 4 contains the effective wordline time constants for various wordline lengths. As shown in the table, the wordline length cannot exceed 512 memory bits (512 digitlines) if the wordline time constant is to remain under four nanoseconds.

Please amend paragraph number **[0108]** as follows:

**[0108]** Folded Array ~~Architecture.~~ Architecture: The folded array architecture depicted in FIG. 39 is the standard architecture used in today's modern DRAM designs. The folded architecture is constructed with multiple array cores separated by strips of sense amplifiers and either row decode blocks or wordline stitching regions. Unlike the open digitline architecture which uses $6F^2$ memory bit cell pairs, the folded array core uses $8F^2$ memory bit cell pairs. Modern array cores include 262,144 ($2^{18}$) functionally addressable memory bits arranged in 532 rows and 1,044 digitlines. The 532 rows consist of 512 actual wordlines, 4 redundant wordlines and 16 dummy wordlines. Each row (wordline) connects to memory bit transistors on alternating digitlines. The 1,044 digitlines consist of 1,024 actual digitlines (512 columns), 16 redundant digitlines (8 columns) and 4 dummy digitlines. As discussed earlier, photolithography limitations necessitate the use of the dummy wordlines and digitlines. These photo problems are very pronounced at the edges of large repetitive structures such as the array core.

Please amend paragraph number **[0112]** as follows:

**[0112]** Similar to the open digitline architecture, digitline length for the folded digitline architecture is again limited by power dissipation and minimum cell to digitline capacitance ratio. For the 256 Mbit generation, digitlines are restricted from having connection to more than 256 cells (128 memory bit pairs). The analysis to arrive at this quantity is similar to that for the open digitline architecture. Refer back to Table 3 to view the calculated results of power dissipation versus digitline length for a 256 Mbit DRAM in 8K refresh. Wordline length is again limited by the maximum allowable RC time constant of the wordline. Contrary to an open

digitline architecture in which each wordline connects to memory bits on each digitline, the wordlines in a folded digitline architecture only connect to memory bits on alternating digitlines. Therefore, a wordline can ~~cross1,024~~ cross 1,024 digitlines while only connecting to 512 memory bit transistors. The wordlines will have twice the overall resistance, but only slightly more capacitance since the wordlines run over field oxide on alternating digitlines. Table 8 contains the effective wordline time constants for various wordline lengths for a folded array core. For a wordline connected to N memory bits, the total resistance and capacitance using 0.3$\mu$m polysilicon are $Rwl = 2 \cdot N \cdot Pwl \div 0.3\mu m$ ohms and $Cwl = Cw8 \cdot N$ Farads, respectively. As shown in Table 8, the wordline length cannot exceed 512 memory bits (1,024 digitlines) for the wordline time constant to remain under four nanoseconds. Although the wordline connects to only 512 memory bits, it is two times longer (1,024 digitlines) than wordlines in open digitline array cores. The folded digitline architecture, therefore, requires half as many row decode blocks or wordline stitching regions as the open digitline architecture.

Please amend paragraph number **[0117]** as follows:

**[0117]** Advanced ~~Bilevel~~ Bi-level DRAM ~~Architecture.~~ Architecture: The present invention provides a novel advanced architecture for use on future ~~large scale~~ large-scale DRAMs. A 32 Mbit memory block is described with this new architecture for use in a 256 Mbit DRAM. The results achieved with the new architecture are compared to those obtained for the open digitline and folded digitline architectures described above.

Please amend paragraph number **[0118]** as follows:

**[0118]** The ~~bilevel~~ bi-level digitline architecture is an innovation which has created a new DRAM array configuration -- one that allows the use of $6F^2$ memory bits in an otherwise folded digitline array configuration.

Please amend paragraph number **[0119]** as follows:

**[0119]** $6F^2$ memory cells are a byproduct of cross-point style (open digitline) array blocks. Cross-point style array blocks require that every wordline connect to memory bit transistors on every digitline, precluding the formation of digitline pairs. Yet, digitline pairs (columns) remain an essential element in folded digitline type operation. Digitline pairs and digitline twisting are important features that provide for good ~~signal to noise~~ signal-to-noise performance. The ~~bilevel~~ bi-level digitline architecture solves the cross-point and digitline pair dilemma through vertical integration. Essentially, two open digitline cross-point array sections 100 are placed side by side as seen in FIG. 41. Digitlines in one array section are designated as true digitlines 106(b) and 104(b) while digitlines from the second array section are designated as complement digitlines 104(a) and 106(a). An additional conductor is added to the DRAM process to complete formation of the digitline pairs. The added conductor allows digitlines from each array section to route across the other array section -- both true and complement digitlines being vertically aligned. At the juncture 108 between each section, the true and complement signals are vertically twisted. This twisting allows the true digitline to connect to memory bits in one array section and the complement digitline to connect to memory bits in the other array section. The twisting concept is illustrated in FIG. 42.

Please amend paragraph number **[0120]** as follows:

**[0120]** To improve ~~signal to noise~~ signal-to-noise characteristics of this design, the single twist region is replaced by three twist regions as illustrated in FIG. 43. An added benefit to multiple twist regions is that only half of the digitline pairs actually twist within each region, thus making room in the twist region for each twist to occur. The twist regions are equally spaced at the 25%, 50%, and 75% marks in the overall array. Assuming that even digitline pairs twist at the 50% mark, then odd digitlines twist at the 25% and 75% marks. Each component of a digitline pair, true and complement, spends half of its overall length on the bottom conductor connecting to memory bits and half of its length on the top conductor. This characteristic balances the capacitance and the number of memory bits associated with each digitline.

Furthermore, the triple twisting scheme guarantees that the noise terms are balanced for each digitline, producing excellent ~~signal to noise~~ signal-to-noise performance.

Please amend paragraph number **[0121]** as follows:

**[0121]** A variety of vertical twisting schemes is possible with the ~~bilevel~~ bi-level digitline architecture. As shown in FIGs. 44A-C, each scheme utilizes conductive layers already present in the DRAM process to complete the twist. Vertical twisting is simplified since only half of the digitlines is involved in a given twist region. The final selection of a twisting scheme is based upon available process technology, yield factors and die size consideration.

Please amend paragraph number **[0122]** as follows:

**[0122]** FIG. 44A illustrates one architecture where a multi-level digitline pair having a digit and digit complement line is fabricated in a semiconductor die. The two digitlines are separated into multiple segments by a vertical twist juncture 108 which allows one segment of a digitline to be vertically located below a segment of another digitline on one side of the twist while having the vertical locations reversed on the other side of the vertical twist. The multi-level digitline pair is preferably fabricated using three levels of the memory die. Two of the levels are used for the digitlines 104(a), 104(b), ~~and~~ 106(a), and 106(b) while the third level is used for conductive lines 105(a) and 105(b). It will be appreciated that the third level is not required, but the horizontal area requirement will be increased by locating lines 105(a) and 105(b) on the second level. The multiple levels are connected via vertical conductive connections 107. Connections 107 can be any conductive material and are preferably fabricated as either a polysilicon plug or a metal plug. Memory cells are coupled to a digitline when that digitline is located vertically below the other digitline. That is, digitlines 104(b) and 106(a) are coupled to memory cells.

Please amend paragraph number **[0123]** as follows:

**[0123]** FIG. 44B illustrates an architecture where the lines 105(a) and 105(b) are fabricated on two separate levels, thereby requiring four levels of the semiconductor. This architecture reduces the horizon area requirements. FIG. 44C illustrates a memory where the conductive lines 105(a) and 105(b) are fabricated above the digitline pair.

Please amend paragraph number **[0124]** as follows:

**[0124]** The architectures of FIGs. 44 A, B and C include one vertical twist and are presented to illustrate different possible fabrication techniques. It will be appreciated that the memory can be designed with multiple vertical twists per digitline pair such that, in total, half of each digitline is coupled to memory cells. Further, the digitlines, conductive lines 105(a) and 105(b), and connections 107 can be fabricated using available conductive materials including, but not limited to, metal, polysilicon and doped regions in the substrate.

Please amend paragraph number **[0125]** as follows:

**[0125]** To further advance the ~~bilevel~~ bi-level digitline architecture concept, its $6F^2$ memory bit was modified to improve yield. Shown in arrayed form in FIG. 45, the ~~'plaid'~~ "plaid" memory bit is constructed using long parallel strips of active area 110 vertically separated by traditional field oxide isolation. Wordlines 116 run perpendicular to the active area in straight strips of polysilicon. Plaid memory bits are again constructed in pairs that share a common contact 122 to the digitline 104 or 106. Isolation gates 118 (transistors) formed with additional polysilicon strips provide horizontal isolation between memory bits. Isolation is obtained from these gates by permanently connecting the isolation gate polysilicon to either a ground or negative potential. The use of isolation gates in this memory bit design eliminates one and two dimensional encroachment problems associated with normal isolation processes. Furthermore, many photolithography problems are eliminated from the DRAM process as a result of the straight, simple design of both the active area and polysilicon in the memory bit. The "plaid" designation for this memory bit is derived from the similarity between an array of memory bits

and tartan fabric -- very apparent in a color array plot. Isolation transistors having a gate connected to a bias potential have been used for isolation in pitch cells in prior memory devices. Isolation gates, however, have not been used in a memory array.

Please amend paragraph number **[0126]** as follows:

**[0126]** In the ~~bilevel~~ bi-level and folded digitline architectures, both true and complement digitlines exist in the same array core. Accordingly, the sense amplifier block needs only one sense amplifier for every two digitline pairs. For the folded digitline architecture, this yields one sense amplifier for every four metal1 digitlines -- quarter pitch. The ~~bilevel~~ bi-level digitline architecture that uses vertical digitline stacking needs one sense amplifier for every two metal1 digitlines -- half pitch. Sense amplifier layout is, therefore, more difficult for ~~bilevel~~ bi-level than folded designs. The triple metal DRAM process needed for ~~bilevel~~ bi-level architectures concurrently enables and simplifies sense amplifier layout. Metal1 is used for lower level digitlines and local routing within the sense amplifiers and row decoders. Metal2 is available for upper level digitlines and column select signal routing through the sense amplifiers. Metal3 can, therefore, be used for column select routing across the arrays and control and power routing through the sense amplifiers. The function of metal2 and metal3 can easily be swapped in the sense amplifier block depending upon layout preferences and design objectives.

Please amend paragraph number **[0127]** as follows:

**[0127]** Wordline pitch is effectively relaxed for the plaid $6F^2$ memory bit used in the ~~bilevel~~ bi-level digitline architecture. The memory bit is still built using the minimum process feature size of 0.3$\mu$m. The relaxed wordline pitch stems from structural differences between a folded digitline memory bit and an open digitline or plaid memory bit. There are essentially four wordlines running across each folded digitline memory bit pair compared to two wordlines that run across each open digitline or plaid memory bit pair. Although the plaid memory bit is 25% shorter than a folded memory bit (3 features versus 4 features), it also has half as many wordlines, effectively reducing the wordline pitch. This relaxed wordline pitch makes layout

much easier for the wordline drivers and address decode tree. In fact, both odd and even wordlines can be driven from the same row decoder block, thus eliminating half of the row decoder strips in a given array block. This is an important consideration since the tight wordline pitch for folded digitline designs necessitates separate odd and even row decode strips.

Please amend paragraph number **[0128]** as follows:

**[0128]** The ~~bilevel~~ bi-level digitline array architecture depicted in FIG. 46 is the preferred architecture for tomorrow's ~~large scale~~ large-scale DRAM designs. The ~~bilevel~~ bi-level architecture is constructed with multiple array ~~cores~~ sections 100 separated by strips of sense amplifiers 124 and either row decode blocks or wordline stitching regions. Wordline stitching requires a four metal process while row decode blocks can be implemented in a three metal process. The array cores include 262,144 ($2^{25}$) functionally addressable plaid $6F^2$ memory bits arranged in 532 rows and 524 ~~bilevel~~ bi-level digitline pairs. The 532 rows consist of 512 actual wordlines, 4 redundant wordlines, and 16 dummy wordlines. There are also 267 isolation gates in each array due to the use of plaid memory bits, but since they are accounted for in the wordline pitch, they can be ignored. The 524 ~~bilevel~~ bi-level digitline pairs consist of 512 actual digitline pairs, 8 redundant digitline pairs, and 4 dummy digitline pairs. The term "digitline pair" is used in describing the array core structure since pairing is a natural product of the ~~bilevel~~ bi-level architecture. Each digitline pair consists of one digitline on metal1 and a vertically aligned complementary digitline on metal2.

Please amend paragraph number **[0130]** as follows:

**[0130]** Unlike a folded digitline architecture that needs a local row decode block connected to both sides of an array core, the ~~bilevel~~ bi-level digitline architecture only needs a local row decode block connected to one side of each core. As stated earlier, the relaxed wordline pitch allows both odd and even rows to be driven from the same local row decoder block. This feature helps make the ~~bilevel~~ bi-level digitline architecture more efficient than alternative architectures. A four metal DRAM process allows local row decodes to be replaced

by either stitch regions or local wordline drivers. Either approach could substantially reduce die size. The array core also includes the three twist regions that are necessary for the ~~bilevel~~ bi-level digitline architecture. The twist region is somewhat larger than that used in the folded digitline architecture, due to the complexity of twisting digitlines vertically. The twist regions again constitute a break in the array structure, necessitating the inclusion of dummy wordlines.

Please amend paragraph number **[0131]** as follows:

**[0131]** As with the open digitline and folded digitline architecture, the ~~bilevel~~ bi-level digitline length is limited by power dissipation and minimum cell to digitline capacitance ratio. In the 256 Mbit generation, the digitlines are again restricted from having connection to more than 256 memory bits (128 memory bit pairs). The analysis to arrive at this quantity is the same as that for the open digitline architecture, except that the overall digitline capacitance is higher since the digitline runs equal lengths in metal2 and metal1. The capacitance added by the metal2 component is small compared to the metal1 component since metal2 does not connect to memory bit transistors. Overall, the digitline capacitance increases by about 25 percent compared to an open digitline. The power dissipated during a read or refresh operation is proportional to the digitline capacitance (Cd), the supply voltage (Vcc), the number of active columns (N), and the refresh period (P) and is given as $Pd = Vccx \cdot (N \cdot Vcc\ (Cd + Cc)) \div (2 \cdot P)$ watts. On a 256 Mbit DRAM in 8K refresh there are 32,768 ($2^{15}$) active columns during each read, write, or refresh operation. Active array current and power dissipation for a 256 Mbit DRAM are given in Table 12 for a 90nS refresh period (-5 timing) at various digitline lengths. The budget for active array current is limited to 200mA for this 256 Mbit design. To meet this budget, the digitline cannot exceed a length of 256 memory bits.

Please amend paragraph number **[0132]** as follows:

**[0132]** Wordline length is again limited by the maximum allowable RC time constant of the wordline. The calculation for ~~bilevel~~ bi-level digitline is identical to that performed for open digitline due to the similarity of array core design. These results are given in Table 4

above. Accordingly, the wordline length cannot exceed 512 memory bits (512 ~~bilevel~~ bi-level digitline pairs) if the wordline time constant is to remain under the required four nanosecond limit.

Please amend paragraph number **[0133]** as follows:

**[0133]** Layout of various ~~bilevel~~ bi-level elements was generated to obtain reasonable estimates of pitch cell size. These size estimates allow overall dimensions for a 32 Mbit array block to be calculated. The diagram for a 32 Mbit array block using the ~~bilevel~~ bi-level digitline architecture is shown in FIG. 47. This block requires a total of 128 256Kbit array cores. The 128 array cores are arranged in 16 rows and 8 columns. Each 4 Mbit vertical section consists of 512 wordlines and 8,192 ~~bilevel~~ bi-level digitline pairs (8,192 columns). A total of eight 4 Mbit strips are required to form the complete 32 Mbit block. Sense amplifier blocks are positioned vertically between each 4 Mbit section. Row decode strips are positioned horizontally between every array core. There are only a total of eight row decode strips needed for the sixteen array cores since each row decode contains wordline drivers for both odd and even rows.

Please amend paragraph number **[0134]** as follows:

**[0134]** The 32 Mbit array block shown in FIG. 47 includes pitch cell layout estimates. Overall size for the 32 Mbit block is found by summing the dimensions for each component. As before, $Height32 = (Tr \cdot Hrdec) + (Tdl \cdot Pdl)$ microns where Tr is the number of ~~bilevel~~ bi-level row decoders, Hrdec is the height of each decoder, Tdl is the number of ~~bilevel~~ bi-level digitline pairs including redundant and dummy, and Pdl is the digitline pitch. Also, $Width32 = (Tsa \cdot Wamp) + (Twl \cdot Pwl6) + (Ttwist \cdot Wtwist)$ microns, where Tsa is the number of sense amplifier strips, Wamp is the width of the sense amplifiers, Twl is the total number of wordlines including redundant and dummy, Pwl6 is the wordline pitch for the plaid $6F^2$ memory bit, Ttwist is the total number of twist regions, and Wtwist is the width of the twist regions. Table 13 shows the calculated results for the ~~bilevel~~ bi-level 32 Mbit block shown in FIG. 47. A triple metal process is assumed in these calculations since it requires the use of local row

decoders. Array efficiency for the ~~bilevel~~ bi-level digitline 32 Mbit array block is given as *Efficiency* = $(100 \cdot 2^{25} \cdot Pdl \cdot 2 \cdot Pwl6) \div (Area32)$ percent, which yields 63.1 percent for this design example.

Please amend paragraph number **[0135]** as follows:

**[0135]** With metal4 added to the ~~bilevel~~ bi-level DRAM process, the local row decoder scheme can be replaced by a global or hierarchical row decoder scheme. The addition of a fourth metal to the DRAM process places even greater demands upon process technologists. Regardless, an analysis of 32 Mbit array block size was performed assuming the availability of metal4. The results of the analysis are shown in Tables 14 and 15 for the global and hierarchical row decode schemes. Array efficiency for the 32 Mbit memory block using global and hierarchical row decoding calculates to 74.5 percent and 72.5 percent, respectively.

Please amend paragraph number **[0136]** as follows:

**[0136]** Architectural ~~Comparison.~~ Comparison: Although a straight comparison of DRAM architectures might appear simple, in actual fact it is a very complicated problem. Profit remains the critical test of architectural efficiency and is the true basis for comparison. This in turn requires accurate yield and cost estimates for each alternative. Without these estimates and a thorough understanding of process capabilities, conclusions are elusive and the exercise remains academic. The data necessary to perform the analysis and render a decision also varies from manufacturer to manufacturer. Accordingly, a conclusive comparison of the various array architectures is not possible. Rather, the architectures will be compared in light of the available data. To better facilitate a comparison, the 32 Mbit array block size data is summarized in Table 16 for the open digitline, folded digitline, and ~~bilevel~~ bi-level digitline architectures.

Please amend Table 16 as follows:

**Table 16**

| Architecture | Row Decode | Metals | 32 Mbit Area ($\mu m^2$) | Efficiency (%) |
|---|---|---|---|---|
| Open Digit | Global | 3 | 29,944,350 | 60.5 |
| Open Digit | Hier | 3 | 32,429,565 | 55.9 |
| Folded Digit | Local | 2 | 40,606,720 | 59.5 |
| Folded Digit | Global | 3 | 32,654,160 | 74.0 |
| Folded Digit | Hier | 3 | 34,089,440 | 70.9 |
| ~~Bilevel~~ | Local | 3 | 28,732,296 | 63.1 |
| Bi-levelDigit | Global | 4 | 24,322,632 | 74.5 |
| ~~Bilevel~~ Bi-level | Hier | 4 | 24,980,376 | 72.5 |
| Digit | | | | |
| ~~Bilevel~~ Bi-level | | | | |
| Digit | | | | |

Please amend paragraph number **[0137]** as follows:

**[0137]** From Table 16 it can be concluded that overall die size (32 Mbit Area) is a better metric for comparison than array efficiency. For instance, the triple metal folded digitline design using hierarchical row decodes has an area of 34,089,440$\mu m^2$ and an efficiency of 70.9%. The triple metal ~~bilevel~~ bi-level digitline design with local row decodes has an efficiency of only 63.1%, but an overall area of 28,732,296$\mu m^2$. Array efficiency for the folded digitline is higher, but this is misleading, since the folded digitline yields a die that is 18.6% larger for the same number of conductors. Table 16 also illustrates that the ~~bilevel~~ bi-level digitline architecture always yields the smallest die area, regardless of the configuration. The smallest folded digitline design at 32,654,160$\mu m^2$ and the smallest open digitline design at 29,944,350$\mu m^2$ are still larger than the largest ~~bilevel~~ bi-level digitline design at 28,732,296$\mu m^2$. Also apparent is that the ~~bilevel~~ bi-level and open digitline architectures both need at least three conductors in their

construction. The folded digitline architecture still has a viable design option using only two conductors. The penalty to two conductors is, of course, a much larger die size -- a full 41% larger than the triple metal ~~bilevel~~ bi-level digitline design.

Please amend paragraph number **[0138]** as follows:

**[0138]** ~~Conclusion.~~ Conclusion: A novel ~~bilevel~~ bi-level digitline architecture for use on advanced DRAM designs has been described. The ~~bilevel~~ bi-level digitline architecture achieves significant reductions in die size while maintaining the high ~~signal to noise~~ signal-to-noise performance of traditional folded digitline architectures. The ~~bilevel~~ bi-level digitline uses vertically stacked digitline pairs connected to arrays of $6F^2$ or smaller memory cells. Vertical digitline twisting ensures balanced noise cancellation and equalizes the quantity of memory cells contacting each digitline. DRAM die size reduction results primarily from the use of smaller memory cells in cross-point style arrays and secondarily from efficient pitch cell utilization. Overall, the ~~bilevel~~ bi-level digitline approach presented combines the best characteristics of both folded and open digitline architectures into an efficient new DRAM architecture.